Are Robots Perceived as Good Decision-Makers: A Study Investigating Trust and Preference of Robotic Referees over Humans in Soccer

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

by

Kaustav Das

May 2019

ABSTRACT

The world is becoming increasingly automated, with more tasks and decisions handled by robots and computers. Today, robots have a presence in decision-making fields like sports, healthcare, and education, but the technology is young and may not be trusted enough to replace a human completely. This study focuses on using this technology in sports, in particular soccer, where increasing popularity has placed increased pressure on referees to make accurate decisions. To improve the accuracy of calls, robotic referees with AI and sensors could be introduced within the next decade. The appearance of the robotic agent and accuracy of call impact how much fans trust them compared to humans. Our online study with 222 participants finds that there is a positive correlation between trust and preference for AI, mechanical, humanoid, and human linesmen. However, despite the rapid rise of technology, participants still prefer human linesmen over the robotic linesmen. The appearance of the robotic linesmen create a psychological barrier to people's readiness to accept the technology, while people's familiarity with human linesmen and expectations of emotions they want to experience from watching soccer- tension and unpredictability of "human error"- serve as a value barrier to change.

Keywords

Human-robot interaction, trust, preference, technology, decision-making

BIOGRAPHICAL SKETCH

I am Kaustav Das. I am a Master of Arts student in Design & Environmental Analysis at Cornell University. I have a minor in Information Science, with a concentration in Human-Computer Interaction. I have a Bachelor of Science degree in Mechanical Engineering from Georgia Tech, with a minor in Industrial Design. I am interested in designing and prototyping creative and innovative products for users. Using my CAD, rapid physical prototyping, machining, and industrial design skills, I can research user needs, brainstorm products and scenarios, design and build concepts, and conduct iterative usability testing to refine the user experience. I look forward to using my skills to take up new responsibilities, make new connections, and overcome new challenges.

ACKNOWLEDGEMENTS

I would like to thank my committee advisors - chair Keith Green (from DEA) and minor member Malte Jung (from Information Science) – for their guidance towards scoping the range and focus of my thesis topic and recruiting participants. I would also like to thank my friends at Cornell University for their support and feedback on the design of my study. Finally, I would like to acknowledge my family for their continued support through the whole journey.

TABLE OF CONTENTS

INTRODUCTION

Soccer is the most popular sport in the world with 4 billion fans worldwide [41]. The competitive nature of the sport naturally evokes a wide range of emotions in fans, from euphoria to grief. Fans expect soccer games to run smoothly and fairly. If the referees and assistant referees (linesmen) on the pitch do not abide by their responsibility to make quick, accurate, and fair decisions during the course of play, fans direct their negative emotions to them. Although the speed and agility of the game and scrutiny of referee decisions (e.g. slow-motion replay from multiple camera angles) has increased over time, the human decision-making ability of the referee has remained constant. For example, in the 2010 World Cup, 8% of the decision events were inaccurate. To maintain the integrity of the game, the accuracy must be improved 3[14]. Soccer referees are always required to keep up with play to ensure optimal positioning in making key decisions like freekicks and offsides. The accuracy of referee decisions is often affected by fatigue and the home advantage bias effect [1].

Television match officials (TMOSs), who are officiating experts outside the pitch with access to more data and camera angles, can make more accurate calls than on-pitch referees. Given the increasingly common use of TMOs in sports such as cricket, rugby union, and rugby league, it would be of interest to analyze the accuracy of the video official's decision having watched an incident several times relative to the frequency of accurate decisions made from the first viewing [19]. There is a greater likelihood of video officials being impartial to the nuances of the game such as player and crowd reactions. They are required to make relatively passive judgments from a number of different perspectives compared to on-pitch referees [20]. The introduction of robotic/smart referees and linesmen (assistant referees) would eliminate controversies caused by "human error" that can be introduced by factors like fatigue and bias. To improve the accuracy of these decisions in soccer, new technology has recently been introduced, for example goal-line sensors in 2014 and VAR (Virtual Assistant Referee) in 2018. Additionally, robotic officials are expected to aid in helping improve the accuracy of crucial calls during the game, and top flight soccer matches could be officiated by robot referees and linesmen by 2030 [1].

BACKGROUND

The referee has responsibility for control of players' behavior during competitive football and implementing the rules of the game. To do this, the referee and the two assistant referees are obliged to keep up with play. Both groups display evidence of fatigue towards the end of the game [21]. Krustrup and Bangsbo reported that referees were further away from infringements in the second half than in the first, suggesting that the fall in work rate manifests as true fatigue [22]. When assessing a referee's physical match performance, there is a need to assess the overall match intensity prior to examining the activity profile of the referee in order to gain a further insight into the pattern of the match. Weston et al.'s study [18] shows that the higher the amount of high intensity running completed during the first half, the lower the ability to sustain high-intensity physical activity during the second half, possibly as a consequence of accumulated fatigue.

On the other hand, what appears as referee favoritism may actually be excessive and illegal aggressive behavior by players in teams that are behind in score. These differences may be wrongly interpreted as evidence of bias where within-game events such as goals scored and recent cards issued were not included. Controlling for such pre-game and within-game variables, Buraimo et al. found evidence for home team favoritism induced by crowd pressure in German and English leagues' home teams, as indicated by home teams receiving fewer cards and on average disproportionately ahead in score [15]. Referees, who are appointed to be impartial, can tend to favor the home team by systematically awarding more stoppage time in close matches in which the home team is behind and also favor the home team in decisions to award goals and penalty kicks [15]. Crowd composition affects the size and the direction of the bias, and the crowd's proximity to the field is related to the quality of refereeing accuracy [16]. The referee is then biased to satisfy the majority of the crowd [17].

On-pitch referees see and hear information that a video or robotic official cannot and can also account for the context of the game [20]. However, match referees are less accurate as the decisions became more difficult. In soccer, one of the most difficult calls for a referee to make is whether a player is offside or not. A player is 'offside' if he or she is closer to the goal than the last defender (excluding the goalkeeper) when the ball is passed to them. The offside official rules from the FIFA laws of the game [29] can be difficult both to be understood by spectators and to be put into practice by experienced people such as referees [30]. A linesman or assistant referee, who has the responsibility of judging offside, is prone to making mistakes. Errors made by assistant referees or linesmen in judging offside may often be the result of the relative optical projections of the players on the AR's retina [2]. Regardless of the quality of the linesman, judgement errors are inevitable, owing to the apparent limitations of our perceptual system [2]. Linesmen appear to need some time to perceptually get used to the typical movements of the defenders and attackers around the offside line. Errors in assessing offside seem to be caused primarily by the flash-lag effect (the tendency for observers to misjudge a moving object at a discrete moment defined by a time marker) and not so much by the inappropriate positioning of assistant referees [3]. Given the difficulty of judging offsides, offside calls are the focus of the investigation reported in this paper.

While the introduction of robotic referees and linesmen is expected to improve the accuracy of calls, the design and physical appearance of a robotic referee may influence how much a person trusts and prefers in making certain calls. A robot's appearance is a key factor because it affects people's moral judgments about that robot [4]. The mere appearance of a robot as human-like seems to invite people to treat this robot similarly to the way they treat a human agent [4]. For instance, hand gestures are a powerful way for human communication, with lots of potential applications in the area of human computer interaction [6]. Zlotowski et al. found that although people rated humanlike communication more important for companion robots than humanlike appearance [24], the appearance of a robot can be an important factor modulating a robot's perceived empathy and trustworthiness. A physical form as opposed to a visual system can affect initial trustworthiness of the robot [25], wherein a highly humanlike robot is perceived as less trustworthy and empathic than a more machinelike robot [26]. Given the impact of a robot's appearance on trusting and preferring a robot, this paper considers the physical appearance of the robot as explored later.

People have shown human-robot asymmetry only when making judgments about a mechanical-looking robot, and not a humanoid robot. The display of a mechanical agent may trigger a mental model of robots as more rational, more "utilitarian," and less affected by guilt and social reputation [4]. The aesthetic appearance of robots affects the perceived social reputation, social presence, and sociability of a robot. A higher quality of social interaction with a robot does not always lead to the higher-satisfaction of the services provided by a robot, which provides interesting implications for the design of a robot [23]. The designer of such a robot system must think about how the behavior of the robot should be designed [6] to maximize the level of trust a soccer fan has in the system. Qualitative data has suggested the appearance of the robot highly biased participants' impressions about its capabilities [7]. In a paper discussing challenges of human-robot interaction in terms of operator trust, Donald Norman, emphasized that in order to build trust, we need to humanize robots, providing them with meaningful emotions and personality. His research has shown that "human error" is not typically a function of just the human, but rather a function of the system in question not accurately/ effectively facilitating the user's understanding of how the system works. Norman has also shown that training does not always overcome trust issues arising from poor design [27].

Studies (eg: [8]) have shown that factors related to the robot itself, specifically, its performance and appearance/design, had the greatest association with trust, while environmental factors were moderately associated, and there was little evidence for the effects of human-related and environmental factors [8]. In soccer, that means the nature of the referee has more of an influence on fans' evaluations than the personalities of the fans or the atmosphere in the stadium. Trust is an important component of HRI (human-robot interaction), as illustrated by direct links to outcomes such as team effectiveness and performance [9]. A goal of HRI, therefore, should be to identify ways in which trust can be measured, quantified, and calibrated in these types of interactions [9]. In soccer, factors such as how well players and fans get along, how much they know each other, and their respective personalities and interests, can lead to very different experiences when it comes to trusting a robot [7]. These experiences of trust can also vary depending on when they are measured - before the interaction, during the interaction, after the interaction, and in the long-term [11].

Notwithstanding the push for technology, there is a hanging question of how much these robotic referees will be trusted and accepted by fans. Even if fans trust a robot more than a human, they may still prefer a human referee or linesman because "human error" brings an element of unpredictability and tension to the game there are still fans who may prefer a human referee over a robotic referee because the element of "human error" in tight decisions is part of the spectacle of the game. The referee for many sports is seen as the "villain," as they may make bad calls (or miss them completely) due to just being human and not being able to see everything. The anger at the referee is what makes watching the sport so frustrating, yet so entertaining and rewarding [10].

The algorithms behind artificial and robotic systems have the potential to improve the efficiency and equity of decisions in soccer [5]. Except in trivial cases for decision-making, it is impossible to maximize accuracy and algorithmic fairness at the same time, and impossible simultaneously to satisfy all kinds of fairness. There is a need to consider challenging tradeoffs [12], which in the case of a robotic referee, involves how much fans trust a referee, and which referee they would prefer to make calls This study aims to gage both the variables trust and preference by asking people their judgements of referee calls. There is possibly an interaction between the variables type of referee, trust, and preference. This study will attempt to test the following hypotheses: 1) The AI linesman is the most trusted because participants perceive it makes the most accurate decisions, and 2) There is a positive correlation between trust and preference. The findings will help sports futurologists determine the best design for a robotic referee (for example, if it should replicate the looks of a human or be more mechanical) that strikes a balance between trust and preference in the application of robot agents to refereeing sporting events.

METHODOLOGY

Pre-Design of Online Study

This online study is was designed to determine which visual features and communication methods has the maximum average trust amongst participants. Participants are also asked to rate which linesman they would prefer the most to make offside calls at a soccer field. The objective of this second part is to determine if there is a direct correlation between trust and preference. If there is no significant correlation, the data can be used to find out what creates a discrepancy between trust and preference. The results of this study can be useful for soccer and other sport organizations (like hypothetical company *Ref-Tech* in this study) to understand what kind of robotic design path to pursue when brainstorming and prototyping robotic referees and linesmen in the future.

Before recruiting participants for the final online survey, a pilot experiment was conducted on the online study to evaluate the feasibility, time, and format (N=4). This helped improve the design of the final full-scale study by making it easier for participants to understand the prompts and flow of the online survey. To make sure the participant of the final survey is not confused with terminology, the term “linesman” is introduced at the start when they view the video on offsides, and the term “linesman” is then used consistently throughout the rest of the survey.

Based on feedback from the pilot, one confounding variable that was brought up was how much knowledge participants have of soccer. Because this can range from no knowledge to very good knowledge, this information was obtained from the participants to explore if their general knowledge of soccer had any influence on what type of linesman they preferred. For participants that have no or little knowledge of soccer, it may be the first time they are introduced to the concept of offsides. In the pilot, it was difficult for this group of people to follow the 16 scenario clips at full speed and look out for offsides. Hence, the clips were slowed down to 70% running speed to compensate for this ambiguity and make it easier for all the participants to distinguish for themselves if a scenario is offside or not. Participants in the pilot study also suggested for controlling for other variables concerning the linesman like the size, position, and speed of reaction. So, for all scenarios, the linesmen were scaled to a similar proportion, positioned to the left on the side of the field, and adjusted to have a similar reaction speed (~0.2 seconds) if making an offside call.

Offsides and Linesmen

Since offside calls are one of the most difficult calls to make during a soccer match by assistant referees (linesmen) [29], this study decided to focus on that aspect of the game. Alternative ways of judging offside, such as off-line analysis of video images taken from an adequate observation point, can help improve the accuracy of offside calls. Smart technology has the ability to improve a soccer linesman's decision-making process. One hypothesis is that technology will improve the fan experience by making quicker and more accurate decisions which fans can trust. Unlike a human referee, a smart or robotic referee is not prone to fatigue during a game or being biased by the home support fans.

The design and visual characteristics of this referee (as fans will view them) is an open question considered in this paper. A robot can be composed of purely mechanical parts, resemble a humanoid, or simply be a "black screen" AI system. The visual cues and communication methods employed by a referee influences how much a spectator (on TV or on the field) trusts the referee and how much he or she would prefer to have that referee make calls during a soccer match. In this study, the two dependent variables to operationalize a fan's experience of a soccer game are trust and preference. The study is formulated as an online survey, where participants assess trust and preference of different referees based on their appearance and calls.

Design of Online Study

The online survey begins by presenting a hypothetical situation to the participant, where a start-up company *Ref-Tech* is trying to determine what kind of robot to use for assistant referees, and if introducing them is the correct move. The results from the survey would hypothetically help *Ref-Tech* make these decisions. This is also where participants are introduced to the terminology of "linesmen", and this is used consistently through the rest of the study. The study then proceeds to show the participants a 42-second video clip about how offsides work, which they can skip if they are familiar with the rule. To make sure they understand the offside rule, participants are then shown two clips (one offside scenario, one onside/ non-offside scenario) and are asked to judge whether it was offside or not. If they make the wrong call, they are shown a modified clip of the correct call in that scenario. This helps prepare participants for the main core of the study.

The core of the study requires participants to view clips of offside calls (similar to the two test scenarios) being made by linesmen of different appearances, and then judge how much they trust the calls. The clips are 11-17 seconds long, and the calls are made by 4 different assistant referees or linesmen. These 4 agents are derived from a similar study by Malle et al. [4] that investigated the impact of the action and appearance of a robot on people's judgements and human-robot (HR) asymmetry. An iterative process was used to develop narrative illustrations based on character and feature development, gathering reference materials featuring robot, human, and AI forms from science fiction, movies, and popular press to arrive at four agent characters: human, humanoid robot, mechanical robot, and AI [4]. Malle et al. first made a selection of initial sketches of agent characters and then developed them with the design goal of simplicity [4].

Figure 1: The four linesmen shown in the clips (left to right: AI, mechanical, humanoid, human) in offside call positions

For this study, appropriate versions of the four agents were found online to suit the soccer setting (see Figure 1): AI blank screen, mechanical robot arm, humanoid robot, and human. Research has shown that facial features, gaze, height, gender, voice, trajectory design, and even proximity to human partners all play a role in how humans respond to robots [32-36]. However, no comprehensive theory predicts when appearance matters, which aspects of appearance matter, and for which psychological or behavior responses it matters when it comes to people *trusting* robot's actions. Thus, accumulating systematic empirical research is key to understanding this relationship.

In the video, each linesman is positioned on the side of the pitch and makes offside calls. The participant watches a total of 16 clips of offside calls, with four calls being made by each of the four linesmen. Of these four offside calls made by each linesman, there is one correct offside call, one wrong offside call, one correct non-offside call, and one wrong non-offside call. These four decision outcomes represent respectively a hit, false alarm, correct reject, and miss. Offside calls are made by the linesman displaying or lifting a red and yellow chequered flag. The order of the type of call and type of linesman is randomized. Hence, the two independent variables in this 4×4 study are "type of call" and "type of linesman," and the dependent variable is trust in the linesman's call. Table 1 shows the combination of independent variables for each of the 16 scenarios.

Table 1: Type of linesman and type of call for each scenario

	Correct offside call	Wrong offside call	Correct onside call	Wrong onside call
AI linesman	Scenario 1	Scenario 11	Scenario 14	Scenario 6
Mechanical linesman	Scenario 5	Scenario 10	Scenario 12	Scenario 3
Humanoid linesman	Scenario 15	Scenario 7	Scenario 4	Scenario 9
Human linesman	Scenario 16	Scenario 2	Scenario 8	Scenario 13

The 11-17-second-long video clips of the 16 scenarios were obtained from a library of offside clips used for testing purposes by the Professional Referee Organization [37]. For some clips from this library, it was relatively easier to judge offsides than others, so the difficulty of these clips was distributed evenly amongst the 4 linesman agents in the videos of the 16 scenarios in the study. The participants were allowed to replay the clips multiple times before finishing the Likert scale ratings. For each of the 16 clips of offside calls, the participant is asked to rate how much they trust the call (see Figure 3). They do so with an empirically based scale developed to measure trust in automated systems and human-machine systems based on performed cluster analysis [31]. This is a Likert scale with the following 12 items: *The system is deceptive, The system behaves in an underhanded manner, I am suspicious of the system's intent, action, or inputs, I am wary of the system, The system's action will have a harmful or injurious outcome, I am confident in the system, The system provides security, The system has integrity, The system is dependable, The system is reliable, I can trust the system, I am familiar with the system.* Each item allows the participant to rate on a scale of 1-7. The data from this scale can be used to analyze how much and why the participants trust a certain type of linesman more than another.

At the end of watching the 16 clips, the participant is asked to rank the four linesmen in order, from the linesman they trust the most to the linesman they trust they least. The participant is then asked to rank the linesmen in terms of how much they would prefer to have them on the field making offside calls. There is also a text box for the participant to comment on the reason behind their choice for their most preferred linesman. The text box aims to capture original data for the two dependent variables in this study: trust and preference. As part of the final background section, which records the participants' age and gender, participants are also asked to rate how much knowledge of soccer they have. This latter information is used to check for potential correlations between knowledge about soccer and the type of referee that is preferred.

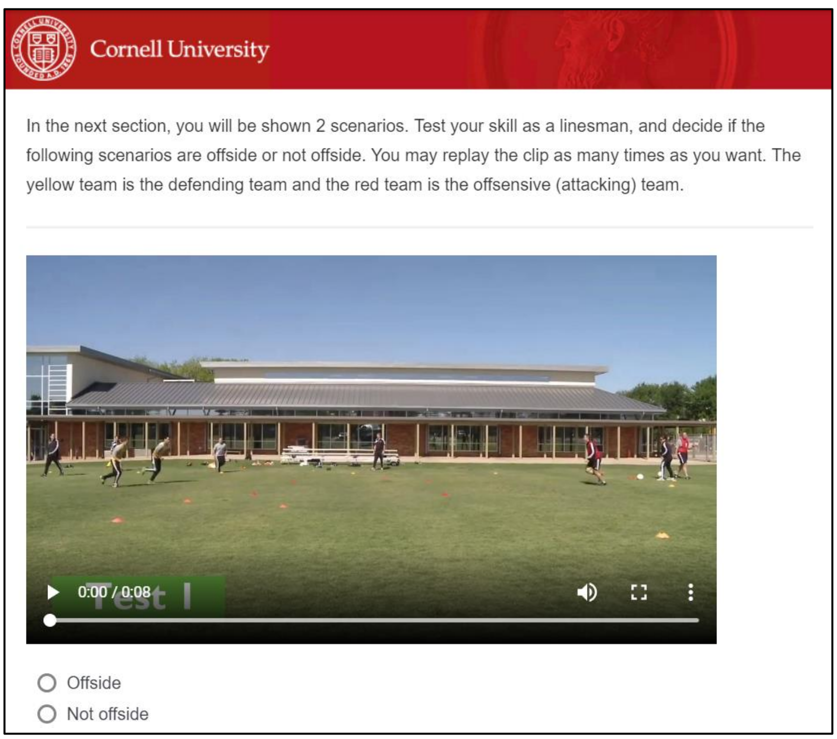
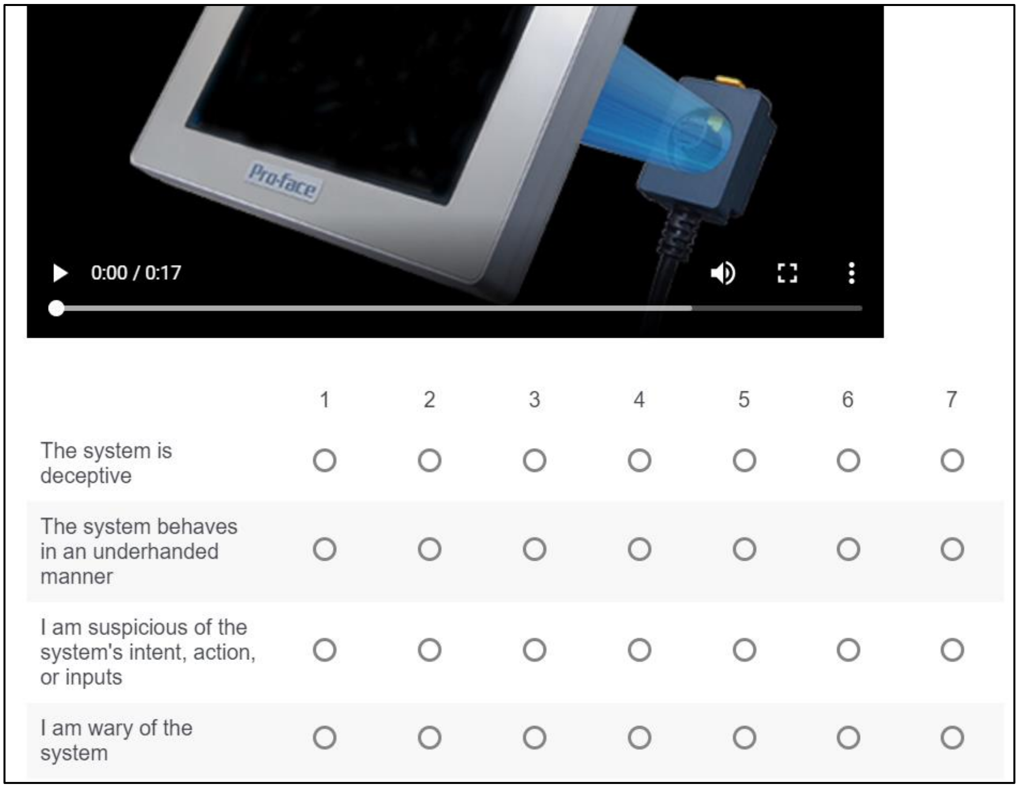


Figure 2: Snapshot of test scenario from study



	1	2	3	4	5	6	7
The system is deceptive	○	○	○	○	○	○	○
The system behaves in an underhanded manner	○	○	○	○	○	○	○
I am suspicious of the system's intent, action, or inputs	○	○	○	○	○	○	○
I am wary of the system	○	○	○	○	○	○	○

Figure 3: Snapshot of Likert scale under a scenario video clip

RESULTS

The study was set up as an external online survey on Cornell University's SONA system to recruit participants. The participants were able to access a link to Qualtrics, where the survey was originally created. A total of 230 people took the study, but 8 of these were omitted for having incomplete data, resulting in a sample size of 222. Of the 222 participants, 155 were female, 66 were male, and 1 chose not to disclose their gender. Because the recruitment was done with the Cornell University system, the majority of participants were college students. The age demographics of the participants read as follows: 72 participants under the age of 20, 145 participants aged 20-29, 3 participants aged 30-39, and 1 participant aged 50-59.

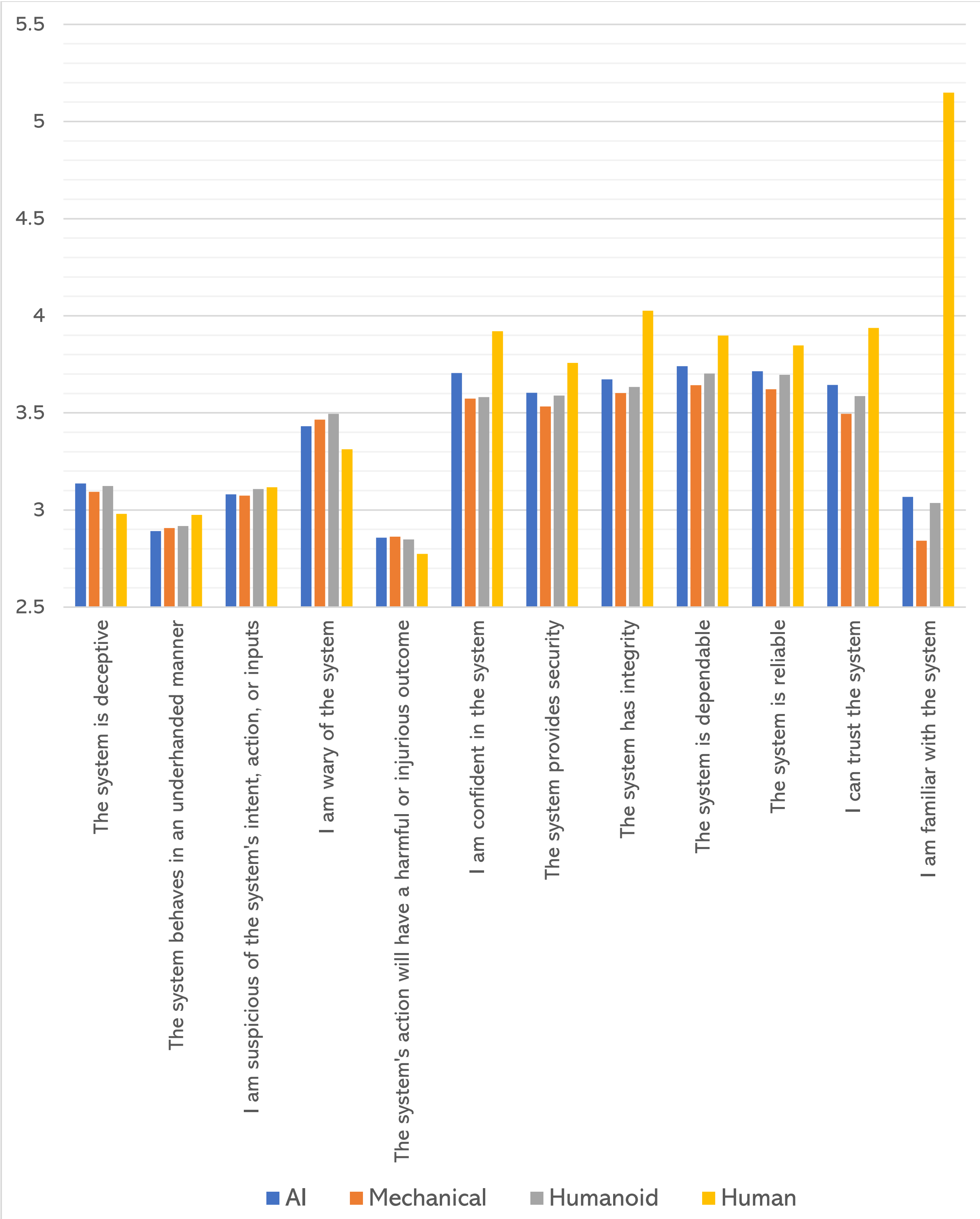


Figure 4: Mean Likert scale ratings of human-robot (HR) trust items for the 4 linesman agents

The first dependent variable that is measured in this study is trust. The empirically developed scale to measure trust between humans and automated systems [31] like robots, which is used to measure trust in this study, has 12 items in total. Each of the items is measured on a discrete scale of 1 to 7. It is important to note that items 1-5 measure trust using words with a negative connotation: “deceptive,” “underhanded,” “suspicious,” “wary,” and “harmful”. Items 6-12 measure trust using words with a positive connotation: “confident,” “security,” “integrity,” “dependable,” “reliable,” “trust,” and “familiar”. The inference from this is that the lower the scores are for items 1-5, the higher the trust a participant has in a linesman’s call. Conversely, the higher the scores are for items 6-12, the higher the trust a participant has in a linesman’s call.

For items 1-5, the human linesman had the lowest score for 3 of the items. For items 6-12, the human linesman had the highest score for all the items. This suggests that on average, participants trusted the human linesman the most. This nullifies the first hypothesis that the AI linesman would be the most trusted because of perceived higher accuracy amongst participants. The data collected from the 222 participants’ responses reported that the average scores for each of the items 1-5 were all lower than the average scores for each of the items 6-12 (see Figure 4). Because correct and incorrect calls were evenly distributed amongst the 16 scenarios to all participants, this reports that there were more instances of participants trusting a wrong call than participants not trusting a correct call.

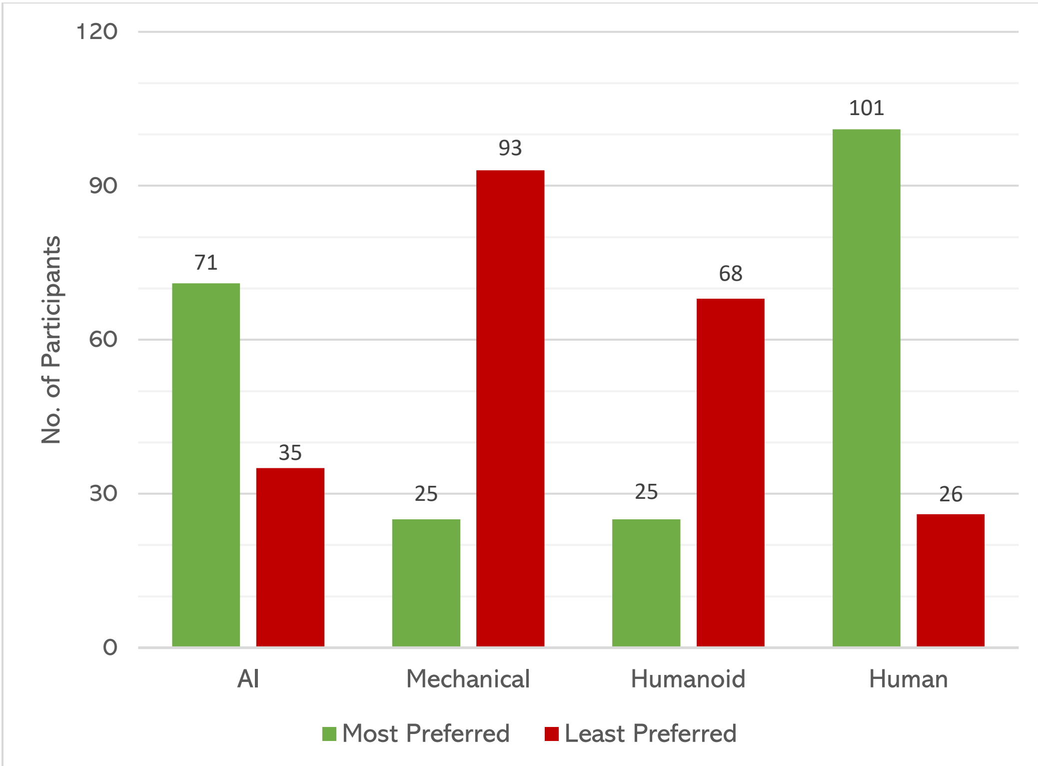


Figure 5: Most and least preferred linesman choices

It is important to note how close the between-group scores are between linesmen for each item. Using standard deviation to measure how spread the values are for each item, item 12 ("I am familiar with the system") is found to have the highest spread (M=3.524, SD=1.899). Figure 3 presents visually that there is very little variance between the average scores for items 1-5 (top 5 items in Figure 4). There is greater variance in scores between different linesman agents for items 6-12 (bottom 7 items in Figure 3), with the AI and human linesmen having higher scores than the mechanical and humanoid linesmen. However, for item 12, the score for the human linesman has a much bigger difference between the scores for the other linesmen as compared to the other items on the scale. This result suggests that familiarity with human linesmen in soccer is the main reason participants trust human linesmen more than humanoid, mechanical, or AI linesmen.

The second dependent variable measured in this study is preference. After completing the trust ratings for all the 16 scenarios, the participants rank the AI linesman, mechanical linesman, humanoid linesman, and human linesman from 1 to 4 in terms of how much they would prefer to have them be the linesman on the field making offside calls. The most preferred (rank 1) choice was the AI linesman for 71 participants, mechanical linesman for 25 participants, humanoid linesman for 25 participants, and human linesman for 101 participants (see Figure 5). The least preferred (rank 4) choice was the AI linesman for 35 participants, mechanical linesman for 93 participants, humanoid linesman for 68 participants, and human linesman for 26 participants. The most common choice for the most preferred linesman was the human linesman (101), and the most common choice for the least preferred linesman was the mechanical linesman (93).

A reliable scale for HR trust has to be used to determine the correlation with the other dependent variable preference. To test the reliability of the empirically developed scale to measure trust between people and automated systems for this study, the Cronbach alpha value was calculated [45]. The scale ratings of HR trust items were determined to be not reliable enough (α=0.6834) since the Cronbach alpha value was below the acceptable threshold of 0.7. To improve the reliability of the scale when determining the correlation between trust and preference, further analysis was done to see closely the twelve items were related to each other. This was done by creating a correlation matrix (see Table 2) between the twelve items. In the matrix, the top row and left column represent the same items. The left column has been shortened to reflect the main words in the item; for example, “Underhanded” is used to represent the item “The system behaves in an underhanded manner”. If the twelve items were to be divided into two groups- the first five items with words with negative connotation and the last seven items with words with positive connotation- there was positive correlation (cells highlighted in green) for items within groups but negative correlation (cells highlighted in red) for items between groups. This means that the score ratings for items with words having negative connotation (the first 5 items) were not close to the score ratings for items with words having positive connotation (the last 7 items) and vice versa. The first five items had similar score ratings to each other and the last seven items had similar score ratings to each other.

The Cronbach alpha was then computed separately for the first 5 items of the HR trust scale and last seven items of the HR trust scale, The Cronbach alpha value was found to be higher for the last seven items (α=0.9413) than the first five items (α=0.9223). For the purposes of calculating the correlation coefficient between trust and preference, the last seven items in the HR trust Likert scale were used because they were found to be more reliable than the first five items based on Cronbach alpha value. Before computing the correlation coefficient between the two dependent variables trust and preference for the 4 linesman agents in the study, both variables were first brought down to a common normalized scale (ranging from 0-1). The final seven items of the HR trust Likert scale *(I am confident in the system, The system provides security, The system has integrity, The system is dependable, The system is reliable, I can trust the system, I am familiar with the system*) were used for trust measure purposes in correlation. These items ranged from 1-7. For each participant, the mean trust score of these last seven items was taken. This score was then divided by 7 to obtain the normalized trust score between 0 and 1, resulting in a discrete range for trust from 0.1429 to 1, with equal intervals of 0.0204. This is represented by the linear black dashed line on Figure 5.

Since preference was measured by ranking (with 1 denoting highest preference), the number was reversed for correlation purposes to denote preference “weights”. So, the value 4 represented highest preference, 3 represented 2nd highest preference, 2 represented 3rd highest preference, and 1 represented lowest preference. To normalize the preference weight scores from 0-1 for each participant, these values were divided by 4. This provided the following normalized preference scores: 1 for highest preference, 0.75 for second highest preference, 0.5 for third highest preference, and 0.25 for lowest preference.

Table 2: Correlation matrix of items in human-robot (HR) scale

	The system is deceptive	The system behaves in an under-handed manner	I am suspi-cious of the system's intent/ action/ inputs	I am wary of the system	The system's outcome will have harmful/ injurious outcome	I am confident in the system	The system provides security	The system has integrity	The system is depen-dable	The system is reliable	I can trust the system	I am familiar with the system
Deceptive	1											
Underhanded	0.818	1										
Suspicious	0.738	0.785	1									
Wary	0.691	0.686	0.784	1								
Harmful	0.615	0.631	0.647	0.635	1							
Confident	-0.397	-0.335	-0.385	-0.468	-0.296	1						
Security	-0.324	-0.261	-0.326	-0.409	-0.230	0.814	1					
Integrity	-0.341	-0.291	-0.353	-0.410	-0.278	0.764	0.803	1				
Dependable	-0.407	-0.358	-0.400	-0.486	-0.319	0.826	0.810	0.805	1			
Reliable	-0.402	-0.353	-0.393	-0.478	-0.319	0.828	0.800	0.789	0.932	1		
Trust	-0.401	-0.356	-0.412	-0.501	-0.325	0.828	0.789	0.789	0.888	0.907	1	
Familiar	-0.130	-0.071	-0.108	-0.199	-0.109	0.430	0.395	0.442	0.427	0.424	0.490	1

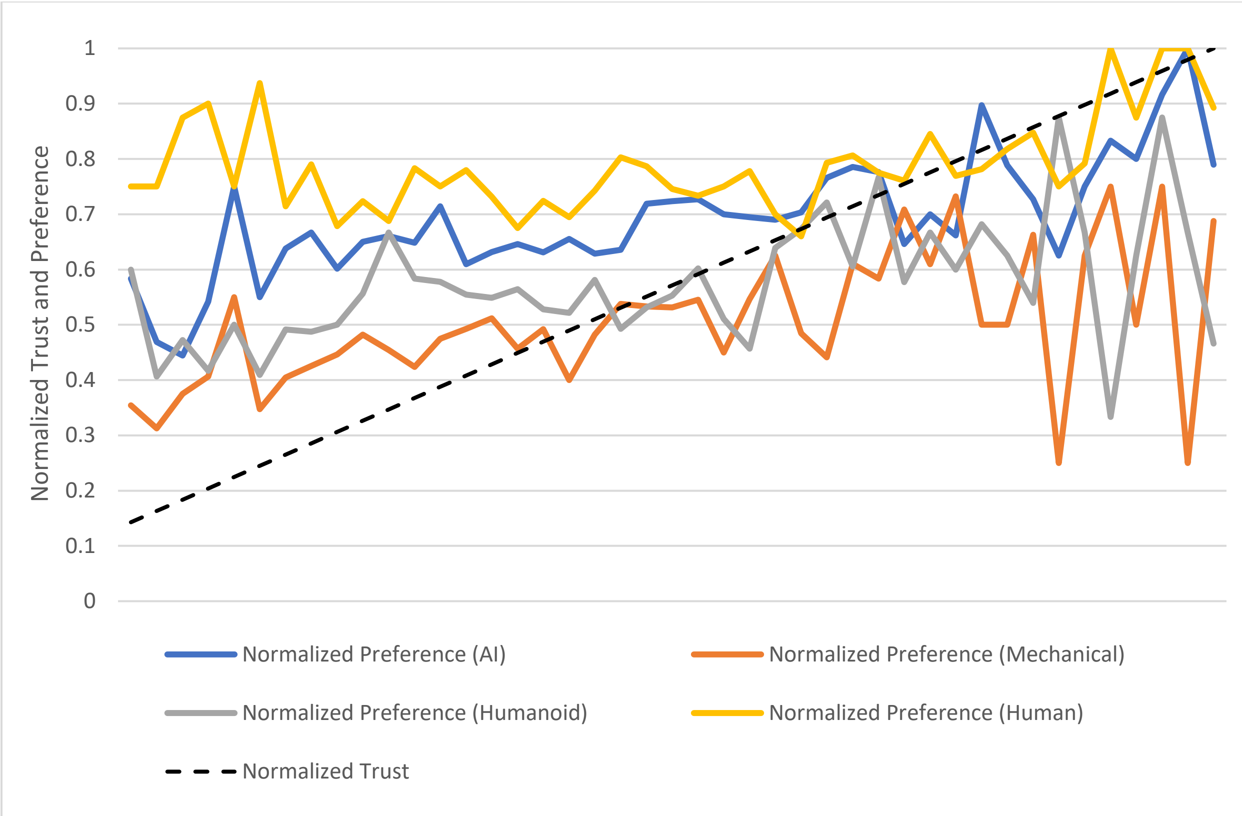


Figure 6: Correlation between trust and preference for the 4 linesmen

Trust was correlated with preference separately for each type of linesman. This means that for studying the correlation between trust and preference for the AI linesman, scenarios 1, 6, 11, and 14 were taken into account when calculating the correlation coefficient (Table 1). Similarly, scenarios 3, 5, 10, and 12 were considered for the mechanical linesman, scenarios 4, 7, 9, and 15 for the humanoid linesman, and scenarios 2, 8, 13, and 16 for the human linesman (Table 1). To study correlation, the 222 participants were separated into groups based on their normalized trust score between 0.1429 and 1. For instance, there were 27 participants with a normalized trust value of 0.1429 for AI (scenarios 1, 6, 11, and 14). The average normalized preference trust scores of this group of 27 students was taken and this provided the first data point for the line graph for normalized preference for the AI linesman (represented by the blue line on Figure 5). The next discrete value for normalized trust range after 0.1429 was 0.1633. For the AI scenarios, there were 8 participants with the normalized trust score of 0.16233. The average normalized AI preference weight scores of these 8 participants was taken to obtain the second data point for the blue line graph for AI normalized preference on Figure 5. This procedure was repeated for all the discrete normalized trust values until the normalized trust value of 1. This helped complete the line graph for normalized preference for the AI linesman (blue line graph on Figure 6).

This whole procedure was then repeated for the other linesmen. For the mechanical linesman (scenarios 3, 5, 10, 12), the average normalized preferences of mechanical linesmen were obtained for each discrete normalized trust participant group and plotted against normalized trust (orange line on Figure 6). The same was performed for the humanoid linesman scenarios (grey line on Figure 6) and human linesman scenarios (yellow line on Figure 6). The combined line graphs on Figure 6 portray how closely the normalized preference scores for each linesman correlated with the normalized trust scores. The relative positions of the 4 normalized preference graphs represents each linesman's average rank. For instance, the normalized preference of human linesmen (M=0.767, SD=0.259) is above the other line graphs and this reflects the fact that the human linesman was the most common choice for most preferred linesman (Figure 5). The second choice for most preferred linesman was the AI linesman; the normalized preference graph for AI linesmen (M=0.569, SD=0.269) was the second highest on Figure 6. Below this was the humanoid linesman's normalized preference (M=0.556, SD=0.252) line graph and the lowest was the mechanical linesman's normalized preference (M=0.498, SD=0.257) line graph. This reflects the fact that the mechanical linesman was the most common choice for least preferred linesman.

For all the linesmen, the normalized preference is more irregular for higher normalized trust values because there are fewer participants per normalized trust score. To better gage the correlation between normalized trust and normalized preference for the 4 linesmen, the correlation coefficient [46] was calculated for the sample of 222 participants for the 4 linesmen. The Pearson's correlation coefficients for the AI linesman (r=0.2049), mechanical linesman (r=0.2539), humanoid linesman (r=0.1207), and human linesman (r=0.1122) all show positive correlation between trust and preference. This confirms the second hypothesis that there is a positive correlation between the two variables trust and preference. This substantiates the general trend, the more a participant in the study trusts the calls made by a linesman, the more likely he or she is to prefer that linesman agent to be the one making offside calls on the pitch. The correlation coefficients for all the 4 linesmen are between 0.1 and 0.3, which means that the strength of positive correlation is small. After ranking the 4 linesmen in the study, the 222 participants commented in text boxes their reasons for their most preferred and least preferred linesmen. It is worth investigating these answers to get a better understanding of how common "trust" is as the reason for a participant preferring a particular linesman agent in the study. The distribution of reasons given by the sample for most and least preferred linesman choices (Figure 7) would give a good qualitative indication of people's impressions of calls made by the linesmen.

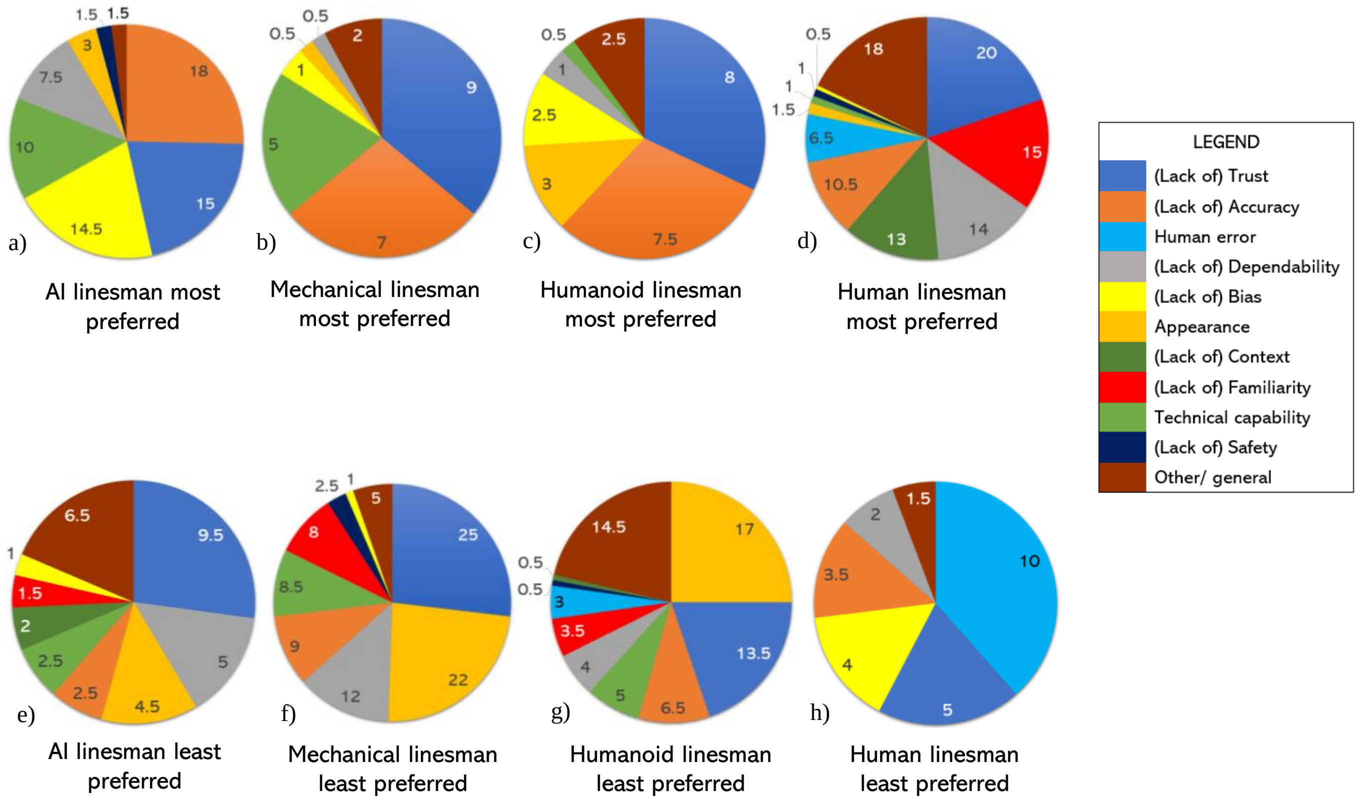


Figure 7: Distribution of reasons provided for a) AI linesman being most preferred choice, b) Mechanical linesman being most preferred choice, c) Humanoid linesman being most preferred choice, d) Human linesman being most preferred choice, e) AI linesman being least preferred choice, f) Mechanical linesman being least preferred choice, g) Humanoid linesman being least preferred choice, h) Human linesman being least preferred choice

The 222 participants were analyzed in groups based on their choices of most preferred linesman (Figure 7 a-d) and choices of least preferred linesman (Figure 7 e-h). Each participant will be part of 2 groups based on their choices. For example, if a participant chose the mechanical linesman as their most preferred linesman and the humanoid linesman as their least preferred choice, the data points for their reasons for these cases will be represented in pie charts 6b and 6g respectively. The reasons listed in the comment section were grouped by cluster analysis based on “key terms.” For example, a comment like “Looks weird and creepy” would fall yield 1 point under the “Appearance” category. Sometimes, participants would have more than one reason for their preference choice. For example, a comment like “Looks potentially sketchy and threatening” would 0.5 point each under the “Appearance” and “Lack of Safety” category. These points are represented on the pie charts for each of the 8 preference groups (Figure 7 a-h). The sum of points for each pie chart represents the number of students in the group. For instance, the sum of points in group 6h is 26, which means a total of 26 participants had the human linesman as their least preferred linesman choice (Figure 5).

Some of the categories in the legend in Figure 6 use words from the HR trust scale, like “Trust,” “Dependability,” and “Familiarity”. There are also other categories like “Accuracy”, “Human Error”, “Bias,” “Aesthetics,” “Context,” “Technical Capability,” and “Safety.” If participants only mentioned that a linesman made more correct calls without hinting at trusting the calls, their reasons were placed under the “Accuracy” category instead of the “Accuracy” category. The difference between “Accuracy” and “Human error” is that the latter refers specifically to the general human tendency to make wrong calls rather than the human linesman in the scenarios making wrong calls. The category “Context” refers to the linesman’s ability to take into account other variables in the game rather than to make a purely objective offside call. In Figure 6, a lot of the categories are used in the positive and negative based on whether the linesman is the most preferred or least preferred choice. For example, the category “Bias” would apply for pie charts 6e-6h, but the category for pie charts 6a-6d would be “Lack of Bias”. Other categories like “Human Error” and “Appearance” apply both positively and negatively. For example, “Human error” is a negative when the participant thinks the general human tendency to make mistakes is detrimental to making offside calls, but it is a positive when participants think that the mistakes and unpredictability that comes from “Human Error” is an important part of soccer as a sport. There were also other reasons participants gave for their most and least preferred linesmen (for example, “Security,” “Confidence,” “Integrity,” “Size”) that were included in the “Other/ general” category because they were less commonly used.

Looking at pie charts 6a-6d for the reasons reported for most preferred linesman choices, “Trust” is the top reason for highest preference for 3 linesman agents (mechanical, humanoid, and human linesmen), which helps explain the positive correlation between trust and preference. For the AI linesman, the top reason for it being most preferred is accuracy, which is the second-highest reason for the mechanical and humanoid linesman being most preferred. The AI linesman and mechanical linesman also scored well in the “Technical Capability” category as the reason for them being most preferred choices. The second-highest reason for the human linesman being most preferred is familiarity, which reflects the high score for human linesmen in item 12 (“I am familiar with the system”) of the HR trust Likert scale compared to the other linesman agents (Figure 3). “Human Error” got a score of 6.5 out of the 101 participants that chose human linesmen as their most preferred choice. Amongst some of the reasons under the “Other/ general” category for the human linesman being most preferred were “Security”, “Confidence”, and “Ethics”. From pie charts 6e-6h which delineate the reasons for least preferred linesman choices, “Lack of trust” is seen to be the highest reason for the AI linesman and mechanical linesman being least preferred, and the second-highest reason for the humanoid linesman and human linesman being least preferred. For the mechanical and humanoid linesman which had the highest scores for least preferred linesmen, the top reasons were “Lack of Trust” and “Appearance”. The top reason for the human linesman (scoring 10 out of 26) being least preferred was “Human Error”.

After recording the participants' reasons for their most and least preferred linesman choices, participants entered their basic background information before concluding the online survey. As part of this background information section, participants were asked to rate on a Likert scale of 1-10 how much general knowledge (GK) of soccer they have. This gives an indication of how much contact participants have with the sport, be it playing it or watching it as a fan. The highest rated general knowledge score was 5 out of 10 (by 32 participants), with more participants scoring below 5 than above 5 (see Figure 8). Using this data, a stacked bar chart was constructed to observe any trends between general knowledge of soccer and most preferred linesman choices (see Figure 9). The highest percentage ratings for the four linesmen are as follows: 1) AI linesman: 55% at GK=9, 2) Mechanical linesman: 27% at GK=4, 3) Humanoid linesman: 25% at GK=7, 4) Human linesman: 50% at GK=5. While the highest percentage for AI linesman being most preferred occurs at GK=9, there is observable significant trend that people with more general knowledge of soccer have a higher likelihood of preferring the AI linesman the most. Moreover, the percentage of AI linesman being most preferred and human linesman being most preferred are both lower for GK=10 than for GK=0. The confounding variable general knowledge of soccer (GK) does not have any noticeable significant impact on the choices of most preferred linesmen.

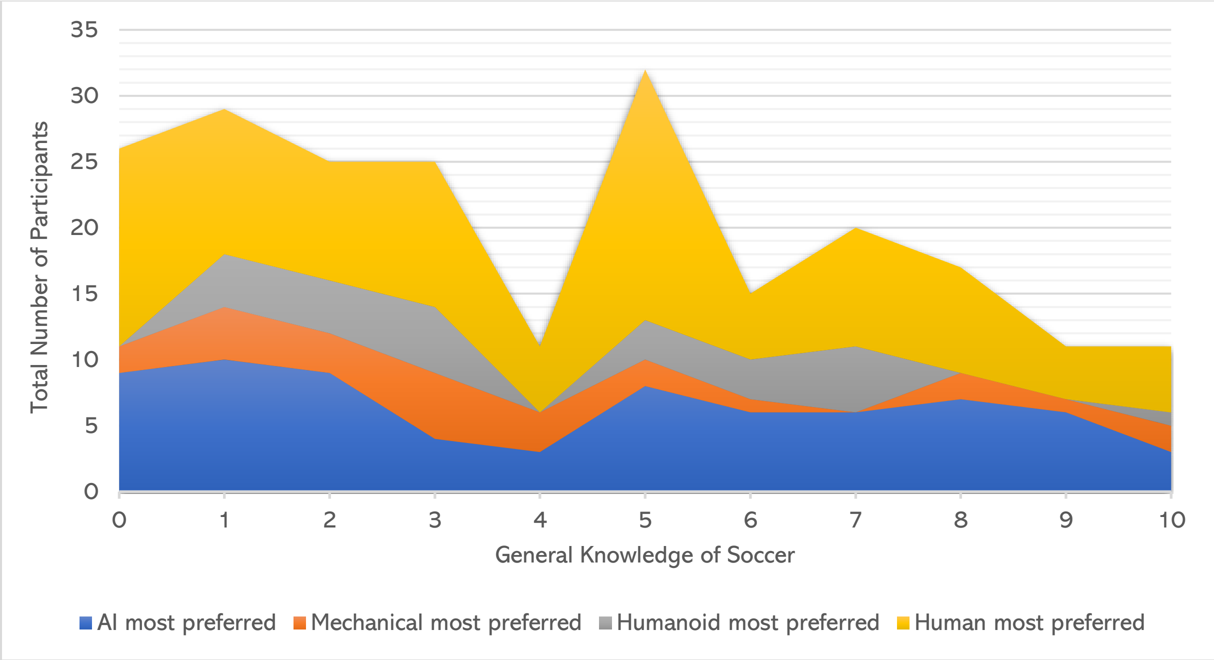


Figure 8: Total number of participants vs general knowledge of soccer

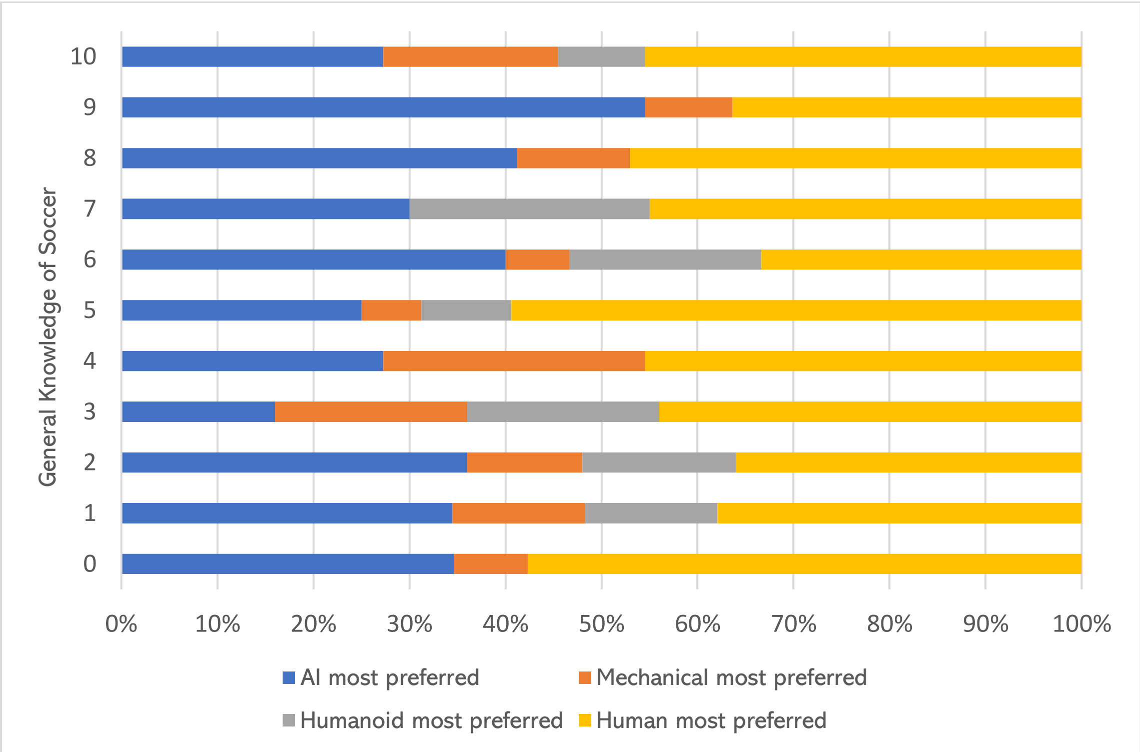


Figure 9: Percentage of most preferred linesman choices vs general knowledge of soccer

DISCUSSION

This study investigated whether people would prefer to have a robotic (the AI blank screen, mechanical robot arm, or humanoid robot) linesman make offside calls in soccer game over a human linesman. The naming convention used was derived from an HRI study that investigated the impact of robot appearance on people's moral judgements [4]. Despite the prediction by some experts that robotic referees and linesmen will officiate soccer games by the year 2030 [1], human linesmen were more often the most preferred linesman choice in this study. The HR Likert scale results convey that the most significant difference in scores lies in item 12 (*I am familiar with the system*). At present day, offside calls amongst other calls are made by human linesmen. Despite the recent introduction of technology like VAR, these human linesmen are still the first point of decision-making, so this is the system spectators or viewers are most familiar with.

This familiarity can serve as a functional barrier that causes viewers to resist changes to robotic referees and linesmen in 3 different ways: 1) Usage barrier, 2) Risk barrier, and 3) Value barrier [38]. Robotic linesmen could prove to have a usage barrier because they do not comply with present practices and habits and would need a change of workflow within players in terms of the speed and accuracy of decision-making as well as not being able to argue with the decision. The AI, mechanical, and humanoid linesmen pose physical risks (potentially harmful and unsafe to players and fans), economic risks (high cost of innovation, especially in countries with low GDP), and functional risks (uncertainty of performance and dependability). The robotic referees and linesmen also pose a barrier to the recreational value of sport spectatorship [39] because they disrupt a spectator's expected affective reactions of tension, euphoria, and unpredictability that provide an agreeable, and sometimes, necessary divergence from their everyday lives [40]. There is also the economic risk of unemployment for thousands of linesmen around the world.

The correlation between trust and preference from the study was found to be positive but weak ($0.1<r<0.3$ for the 4 linesman agents). This is because there are reasons other than "Trust" why participants picked a linesman agent as their most preferred choice (Figure 7). The mechanical linesman has "Trust" as the top reason for being the most preferred choice and "Lack of Trust" as the top reason for being the least preferred choice, which explains why it has a higher Pearson's correlation coefficient between trust and preference than the other linesman agents. The second highest reason for the mechanical linesman being the least preferred is "Appearance": many participants commented on the "scary" look, which in turn makes it seem unsafe and technically incapable. "Appearance" also played the biggest part in participants choosing the humanoid linesman as their least preferred choice, since the resemblance to a human seemed more "creepy" than realistic. In comparison, the simplistic and clean look of the AI linesman may have contributed to more participants preferring it to the other robotic linesmen. The term AI also made some participants assume it is accurate, objective, and unbiased. Human linesmen make more subjective decisions, but they were still the most popular choice for most preferred linesman among the four agents. It is interesting to note that the subjectivity of "Human Error" is seen as a boon and a bane by different participants to the value of spectatorship in soccer [39]. The ability of human linesmen to take into account the other variables and activities ("Context") in the environment before making the decision helps add to the subjectivity and unpredictability of the game, but according to other viewers, this can introduce bias which is a detriment to fairness in running a soccer game.

CONCLUSION AND FUTURE WORK

Human linesmen are the most preferred linesman choice compared to the robotic linesman agents (AI, mechanical, and humanoid linesmen) in this study. There was observed to be a positive correlation between trust and preference. The two independent variables, "type of linesman" and "type of call," both have a main effect on the dependent variable trust. If the type of call was a hit or correct reject, the trust score was higher than when the type of call was a false alarm or miss. The type of linesman also had a main effect where the AI linesman and human linesman were more trusted than the mechanical and humanoid linesmen. The human was more trusted and preferred mostly because of participants' familiarity with the system, and because the AI terminology caused many participants to assume it makes unbiased decisions. The appearance of the mechanical and humanoid linesman played a part in lowering how participants' trust and preference in them because they looked unsafe and creepy respectively, which hinders fans' emotional experience of watching a soccer game. The general knowledge (GK) of soccer amongst participants had no significant impact on which linesman they preferred. In the future, GK can be correlated with the cluster analysis of reasons for most/ least preferred linesmen to study if a participant with more GK feel a different way about a particular linesman agent than a participant with low GK.

Of the robotic linesmen, the AI is least human-like. The mechanical and humanoid linesman try to take up the look of a human by mimicking communicative gestures and body features respectively. However, the mechanical and humanoid linesman score lowest in trust and preference. This presents a strong case for human-robot (HR) asymmetry in a player's or fan's interaction with a robotic linesman. The mechanical and humanoid linesmen have a similar appearance to humans but different behavioral and cognitive strengths and limitations [44]. To build on the four linesman agents used in the study, a wider range of robotic agents and combinations of humans with robots could be tested for trust and preference.

Trust is essential for human-robot collaboration and user adoption of autonomous systems [43]. In the field of healthcare for instance, the preferences of patients in regard to paternalism, clarification, and participation are significantly correlated with trust [28, 42]. Similarly, in this study, there was a positive correlation between the variables trust and preference for all four linesman agents. This study found that there are certain variables like “Dependability” and “Familiarity” that have a main effect on both trust and preference of robotic linesmen, but also contributes to the literature regarding robots as decision-makers I other fields like healthcare, education, and transportation. One limitation of the study is that it explores only four linesman agents, so future studies can explore a wider range of robotic agents by combining visual appearance and communication features in different combinations. Finally, it is worth repeating the study in a few years when robotic systems are more familiar to people to learn if there is a difference in trust and preference scores for robots in the decision-making process.

REFERENCES

[1] Setterfield, Justin. "Robot Football Referees and PLAYERS Could Be a Reality Sooner than You Think." *Mirror*, Mirror.co.uk, 16 Feb. 2018, www.mirror.co.uk/sport/football/news/robot-football-referees-linesmen-could-12030671.

[2] Oudejans, R. R., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbrückner, M., & Beek, P. J. (2000). Errors in judging 'offside'in football. *Nature*, *404*(6773), 33.

[3] Helsen, W., Gilis, B., & Weston, M. (2006). Errors in judging "offside" in association football: Test of the optical error versus the perceptual flash-lag hypothesis. *Journal of sports sciences*, *24*(05), 521-528.

[4] Malle, B. F., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016, March). Which robot am I thinking about?: The impact of action and appearance on people's evaluations of a moral robot. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 125-132). IEEE Press.

[5] Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, August). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 797-806). ACM.

[6] Dylla, F., Ferrein, A., Lakemeyer, G., Murray, J., Obst, O., Röfer, T., ... & Wagner, T. (2008). Approaching a formal soccer theory from behaviour specifications in robotic soccer. *WIT Transactions on State-of-the-art in Science and Engineering*, *32*.

[7] Vázquez, M., May, A., Steinfeld, A., & Chen, W. H. (2011, May). A deceptive robot referee in a multiplayer gaming environment. In *Collaboration Technologies and Systems (CTS), 2011 International Conference on* (pp. 204-211). IEEE.

[8] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517-527.

[9] Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y., & Hancock, P. A. (2011, September). A model of human-robot trust: Theoretical model development. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 55, No. 1, pp. 1432-1436). Sage CA: Los Angeles, CA: SAGE Publications.

[10] Goh, J. W., Pavlovich, E., & Collins, Q. (2018). Fencing Referee Bot.

[11] Billings, D. R., Schaefer, K. E., Chen, J. Y., & Hancock, P. A. (2012, March). Human-robot interaction: developing trust in robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 109-110). ACM.

[12] Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2017). Fairness in criminal justice risk assessments: the state of the art. *arXiv preprint arXiv:1703.09207*.

[13] Freedy, A., Weltman, G., Freedy, E., Parasuraman, R., Visser, E. D., & Coeyman, N. (2007). *Adaptive delegation interfaces (adi) for improved situation awareness and reduction of workload in controlling multiple unmanned vehicles (uv)*. PERCEPTRONICS SOLUTIONS INC SHERMAN OAKS CA.

[14] Solomon, A. V., Paik, C., Alhauli, A., & Phan, T. (2011, April). A decision support system for the professional soccer referee in time-sensitive operations. In *2011 IEEE Systems and Information Engineering Design Symposium* (pp. 35-40). IEEE.

[15] Buraimo, B., Forrest, D., & Simmons, R. (2010). The 12th man?: refereeing bias in English and German soccer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(2), 431-449.

[16] Dohmen, T. (2008) The influence of social forces: evidence from the behavior of football referees. Econ. Inq., 46, 411–424.

[17] Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2005) Favoritism under social pressure. Rev. Econ. Statist., 87, 208–216

[18] Weston, M., Castagna, C., Impellizzeri, F. M., Rampinini, E., & Abt, G. (2007). Analysis of physical match performance in English Premier League soccer referees with particular reference to first half and player work rates. *Journal of Science and Medicine in Sport*, *10*(6), 390-397.

[19] Mascarenhas, D. R., Button, C., Hara, D., & Dicks, M. (2009). Physical performance and decision making in association football referees: A naturalistic study.

[20] Nevill AM, Balmer NJ, Williams M. The influence of crowd noise and experience 7 upon refereeing decisions in football. Psychology of Sport and Exercise. 8 2002;3:261-72.

[21] Reilly, T., & Gregson, W. (2006). Special populations: The referee and assistant referee. *Journal of sports sciences*, *24*(07), 795-801.

[22] Krustrup, P., & Bangsbo, J. (2001). Physiological demands of topclass soccer refereeing in relation to physical capacity: Effect of intense intermittent exercise training. Journal of Sports Sciences, 19, 881 – 891.

[23] Kwak, S. S. (2014). The impact of the robot appearance types on social interaction with a robot and service evaluation of a robot. *Archives of Design Research*, *27*(2), 81-93.

[24] K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, and I. Werry, "What is a robot companion - friend, assistant or butler?," in IEEE IRS/RSJ International Conference on Intelligent Robots and Systems, pp. 1488–1493, IEEE Computer Society, 2005

[25] K. E. Schaefer, T. L. Sanders, R. E. Yordon, D. R. Billings, and P. A. Hancock, "Classification of robot form: Factors predicting perceived trustworthiness," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 56, pp. 1548–1552, Sept. 2012.

[26] Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2016). Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy. *Paladyn, Journal of Behavioral Robotics*, *7*(1).

[27] Bruemmer, D., Few, D., Goodrich, M., Norman, D., Sarkar, N., Scholtz, J., ... & Yanco, H. (2004, April). How to trust robots further than we can throw them. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1576-1577). ACM.

[28] Ommen, O., Janssen, C., Neugebauer, E., Bouillon, B., Rehm, K., Rangger, C., ... & Pfaff, H. (2008). Trust, social support and patient type—associations between patients perceived trust, supportive communication and patients preferences in regard to paternalism, clarification and participation of severely injured patients. *Patient Education and Counseling*, *73*(2), 196-204.

[29] FIFA. Laws of the Game 2007/2008. Available: http://www.fifa.com

[30] D'Orazio, T., Leo, M., Spagnolo, P., Mazzeo, P. L., Mosca, N., Nitti, M., & Distante, A. (2009). An investigation into the feasibility of real-time soccer offside detection from a multiple camera system. *IEEE Transactions on Circuits and Systems for Video Technology*, *19*(12), 1804-1818.

[31] Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53-71.

[32] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler, “All robots are not created equal: The design and perception of humanoid robot heads,” in Proceedings of the 4th Conference on Designing Interactive Systems (DIS ’02): Processes, Practices, Methods, and Techniques, 2002, pp. 321–326.

[33] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, “‘If you sound like me, you must be more human’: On the interplay of robot and user features on human-robot acceptance and anthropomorphism,” Proc. 7th ACMIEEE Int. Conf. Hum.-Robot Interact. HRI12, pp. 125–126, Mar. 2012.

[34] J. Forlizzi, “Towards the design and development of future robotic products and systems,” p. 506, Aug. 2007.

[35] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, “Effects of robot motion on human-robot collaboration,” Proc. Tenth Annu. ACMIEEE Int. Conf. Hum.-Robot Interact. HRI 15, pp. 51–58, 2015.

[36] P. J. Hinds, T. L. Roberts, and H. Jones, “Whose job is it anyway? A study of human-robot interaction in a collaborative task,” Hum.- Comput. Interact., vol. 19, no. 1–2, pp. 151–181, Mar. 2004.

[37] Organization, Professional Referee, director. YouTube. YouTube, YouTube, 10 July 2015, www.youtube.com/watch?v=7K_Hl5Y6lSI

[38] Ram, S., & Sheth, J. N. (1989). Consumer resistance to innovations: the marketing problem and its solutions. *Journal of consumer marketing*, *6*(2), 5-14.

[39] Zillmann, D., Bryant, J., & Sapolsky, B. S. (1989). Enjoyment from sports spectatorship. *Sports, games, and play: Social and psychological viewpoints*, *2*, 241-278.

[40] Cashmore, E. (2005). *Making sense of sports*. Routledge.pp. 6

[41] Sawe, Benjamin Elisha. “The Most Popular Sports on the World.” *World Atlas*, 16 Sept. 2016, https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html

[42] Patel, A., Patel, M., Lytle, N., Toro, J. P., Medbery, R. L., Bluestein, S., ... & Lin, E. (2014). Can we become better robot surgeons through simulator practice?. *Surgical endoscopy*, *28*(3), 847-853.

[43] Chen, M., Nikolaidis, S., Soh, H., Hsu, D., & Srinivasa, S. (2018, February). Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 307-315). ACM.

[44] Vollmer, A. L., Schillingmann, L., Rohlfing, K. J., & Wrede, B. (2014, March). Humans and robots in asymmetric interactions. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 506-507). IEEE.

[45] “HOME.” *IDRE Stats*, https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/

[46] Ganti, Akhilesh. “Correlation Coefficient Definition.” *Investopedia*, Investopidia, 31 Mar. 2019, www.investopedia.com/terms/c/correlationcoefficient.asp